

Data Donations for Mapping Risk in Google Search of Health Queries:

A case study of unproven stem cell treatments in SEM

This is a preprint! The final version has been published in the 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (<https://doi.org/10.1109/SSCI47803.2020.9308420>)

Martin Reber*, Tobias D. Krafft†, Roman Krafft‡, Katharina A. Zweig§ and Anna Couturier¶

* Technische Universität Kaiserslautern; Algorithm Accountability Lab; Germany;

† Technische Universität Kaiserslautern; Algorithm Accountability Lab; Germany;

‡ Technische Universität Kaiserslautern; Algorithm Accountability Lab; Germany;

§ Technische Universität Kaiserslautern; Algorithm Accountability Lab; Germany;

¶ University of Edinburgh; Science, Technology and Innovation Studies; Scotland, United Kingdom;

Abstract—On October 1st, 2019 in response to critique from patient advocates and the medical community, Google explicitly prohibited promotion of unproven stem cell and gene therapy treatments on their platform in order to protect users from rising direct-to-consumer marketing of unproven medical interventions. This project aims to record the efficacy of that prohibition as it was enforced and track the impact of Google’s AI-based advertising modalities on end-user results. In particular, this study gives special consideration to the risk potential for vulnerable patient communities navigating health information through Google search. Utilising a crowd-sourced ‘Black Box’ audit with a browser plugin, we captured the continued presence of prohibited and problematic advertisements returned by stem cell-related queries in the months following Google’s ban. In the domain of Search Engine Marketing (SEM), emerging stem cell treatments are situated in a critical juncture between advertisers and potentially vulnerable users with Google Search as an unobserved mediator. Addressing the issues raised by this data collection is of utmost importance in the protection of patient populations online. This project aims to draw attention to the need for transparency and accountability of advertising intermediaries engaged in the targeted promotion of potentially problematic treatments to vulnerable audiences.

1

I. INTRODUCTION

Among all the industries that digitalisation has accelerated or transformed, the domain of *Stem Cell Treatments* offers a

valuable example of progress in need of close inspection. The rapidly developing area of emerging stem cell-derived treatments within medical research is complemented by a growing corollary of direct-to-consumer, internet-based industry that encourages patients and their carers to access treatment options outside traditional local healthcare infrastructure². Often the commercial stem cell treatments offered provide little to no clinical evidence of safety or benefit for the patient [2], [3]. As this industry relies heavily on the internet [4]–[6] and its associated technologies, direct-to-customer marketing via search engines is a channel of choice for communication from business to patient [7] in pharmaceutical domains [8] and an opportunity to claim search terms [9, p. 184]. Thus, search engines, web advertisement and especially the ad exchange [10]–[12] play crucial roles in connecting the participants of the aforementioned market. By reaching into the medical domain, however, these algorithms have the “most immediate and profound personal and social consequences” [4, p.368].

On principle, Google prohibits advertising in some sensitive categories [13] like personal hardship. In response to the return of problematic advertisements onto Google’s search engine result page (SERP) (see figure 1), the company announced in September 2019 to rigorously ban stem cell related experimental medical treatments as of October 1st 2019 [14], [15]. The data collection of this study concentrates on Google as, to our

¹This is a preprint! The final version has been published in the 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (<https://doi.org/10.1109/SSCI47803.2020.9308420>)

²Often this includes *Stem Cell Tourism*; a call to cross geographical or jurisdictional boundaries to receive medical treatment [1].

best knowledge, they were the only web search and advertising provider that specifically banned experimentally or unproven stem cell therapies and publicly announced a crackdown on those ads, knowing that those are an issue on their platform. This announcement was also acknowledged in media which made it a claim worth to investigate [16].

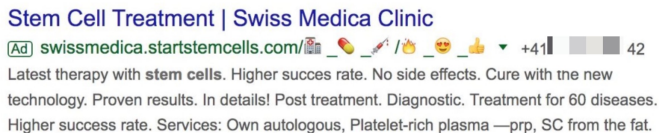


Fig. 1. Single ad of questionable stem cell treatment provider (September 30th, 2019) before the policy change, courtesy of Anna Couturier

Despite their social relevance these platforms are a *Black Box* to society, due to their private, proprietary business model [17], the complexity of their functionality [18]–[21] and the responsiveness to the respective context in which they are applied [22]–[24]. Severe problems are fixed ex post of implementation and often only after it had been pointed out in public discourse and media. Instead of ensuring that the overall viability of a system is increased, platforms seems to address individual problems only. Dörner names this elimination of obvious faults without analysing the fundamental problems “repair service behaviour” [25].

In this paper, we present a browser plugin that was used to assess the success of one such policy update and snapshot of SERP returns. We chose a *Black Box* approach, specifically a *Crowdsourced Audit* [26] to gather user data from volunteers’ browsing experience by sending automated requests to Google’s web search. Participants were assigned to groups according to the disease with which they were affected³ and their home country. Then, the scraped ads and search results were categorised and analysed to assess the risk that web search poses for people affected by diseases like Parkinson’s Disease and Multiple Sclerosis. Thus we apply a concept for algorithmic accountability by striving to investigate the “power structures, biases, and influences that computational artifacts play in society” [27, p.3]. It is our hope that these findings will contribute to a larger civic conversation to hold institutions accountable for their technical creations and demand for explanation, auditability and ultimately access [28].

II. FUNDAMENTALS

This section contextualises this study within the industry that has evolved around stem cell treatments. It explores the benefits gained by this sector from different aspects of digitalisation including data collection, data fusion, profiling and targeting build the basis for targeted advertising of medical treatments.

³This study considers Parkinson’s disease (PD), Multiple Sclerosis (MS) and Diabetes Type I and II because these are diseases for which there is ongoing research into stem-cell-derived treatments. The analysis concentrates on the PD study groups.

A. Stem Cell Treatments: Possibilities and Dangers

This project focuses on the area of stem cell treatments due to their unique positioning within both the areas of emerging medicine and digital health marketing.

Stem cell treatments derived from reprogrammed adult stem cells (‘induced pluripotent stem cells’ or ‘iPS cells’) have received a significant focus within medical research with the promise of personalised therapies [29]. However, as of 2020, the basic characteristics of iPS cells and their capabilities are still far from conventional clinical application. Moreover, outside of specific (largely ‘allogenic’ or tissue donor-reliant) treatments of blood and skin diseases and treatment for corneal damage, there are currently no other proven safe or effective stem-cell-derived treatments for major conditions or diseases [30].

Despite the lack of established treatment protocols, we are witnessing a boom in direct-to-consumer marketing of ‘autologous’ (or obtained from the same individual) stem cell treatments for major diseases and conditions (including Parkinson’s, Multiple Sclerosis, and Diabetes) by private clinics in digital spaces [31]. These clinics and providers present a particular representation of stem cell treatments as established, if restricted, medical interventions available for a price [32]. Moreover, these clinics dedicate a significant portion of their budget to specifically online marketing activities in order to recruit new patients [33].

In the face of the rising profile of direct-to-consumer stem cell treatments online, traditional medical authorities have published strong warnings on the risks for patients associated with undergoing experimental treatments as well as collective calls for stricter regulation for consumers (e.g. [34]–[38]). These calls-to-actions have come on the heels of a number of high-profile cases of private clinic closures after patients suffered gross bodily harm or death after undergoing treatment [39].

Looking beyond the provider and clinic aspects of this digital health interplay, this project interrogates the as-yet unexamined role of Google as a crucial mediator between commercial medical care and risk for patients-as-users. In doing so, we hope to further situate the importance of algorithmic governance and oversight within the conversation of health online.

B. User data as a fuel to online advertising

Online marketing has evolved in its tailoring to individuals through the deduction of personal information beyond what is volunteered by the targeted audience [40]. As a fundamental cornerstone of online advertisement, companies and web-services collect and combine data to derive personally identifiable information (PII) from, at first, neutrally-treated users as well as their usage data in order to create unique profiles [41]–[43]. Sparse individualised data like browsing histories or product ratings are sufficient to de-anonymise users as demonstrated by Narayanan in 2008 [44]. This profiling most likely does not happen by manual categorization, instead it is based on methods of machine learning [45] which are highly opaque.

Profiling, especially when it is based on machine learning models can generate problematic categories and unwanted side-effects that allow discrimination or questionable targeting of users. Angwin et al. showed how Facebook allowed targeting users based on expressed antisemitism [46] or exclusion of users based on race [47]. Furthermore, protected attributes carrying sensitive personal information are often used in the profiling process. This potentially exposes users to discrimination and further restricts the ability to comprehend the reason behind ad choices by the end-user [48]. This is particularly critical if users are unaware of the information being gathered to construct their persona. Indeed, Datta et al. found that users could not review all data that was used by Google to create their profile [48].

C. Targeted advertisement

To fund their operations, search engines often display promotional results alongside their organic ones⁴. They are similar styled as organic results, but marked as advertisement. Once potential customers are profiled and recognized online, they can be targeted with personalised advertising and search results in real time [50], [51]⁵. The use of behavioral targeting in advertisement has been shown to increase click-through rates significantly compared to non-targeting controls in similar users of a distinct audience [53]. Personalised targeting also enhances the efficacy of persuasive marketing strategies and motivates purchases [54]. Google leverages the capabilities of artificial intelligence (AI) or machine learning to enhance ad delivery in many facets such as user profiling and targeting, price negotiation for inventory slots, ad placement and user interaction predictions [55]–[58].

In order to circumvent potential misconduct on the part of advertisers, Google has established a guide to prohibited practices on its *Advertising Policy Help* website [59]. Therein, Google outlaws false statements about qualifications and claims that promise unrealistic results. Research on alcohol ads points out how online advertisement can substitute offline ads in domains where the latter are banned or heavily regulated. It further suggests that this effect is strongest for new products [60]. Search engine marketing (SEM) also allows the “branding” of search terms, consequently associating specific queries with a brand, product or service [9, p. 184]. These two aspects are relevant in the context of emerging medial practices like unproven stem cell treatments that are advertised online despite a ban being effective. In the context of this work, Google’s intervention (or non-intervention) in monitoring misleading content presents compelling questions.

D. Analysing algorithmic ad targeting as a black box

One method to establish Algorithmic Accountability is to conduct a *Black Box analysis*. Black Box analysis is a form of

⁴“organic” denotes unpaid results on the search engine result page that are listed due to their relevance to the search query [49]

⁵“If an ad network is able to accurately target users, we can deduce that the ad network is able to determine user characteristics” [52, p.1].

reverse engineering⁶ whereby an opaque system is scrutinised by analysing observable in- and outputs, deducing the inner mechanics that transform the former into the latter and approximating the inner workings with models. This can be achieved by manipulation and observation of the box [61]. The insights are usually juxtaposed to expectations with respect to certain statistics, norms or standards of stakeholders about how the system is intended to work [24]. This kind of analysis tries to produce a (computational or mathematical) model of an algorithm. To analyse algorithmic decision making systems (ADM systems), scholars have developed different approaches [62], [63] depending on the characteristics of the audit, for example **Scraping Audits** and **Crowdsourced Audits**.

III. STUDY DESIGN

We chose to combine **Scraping Audit** and **Crowdsourced Audit** [62]. In doing so, we recruited volunteers to install a custom-coded browser plugin/addon on their desktop machines. The plugin was designed to regularly scrape the SERP of Google and capture advertisements as well as search results and submit them to a central collection server running at the TU Kaiserslautern, Germany. We also guaranteed a baseline of submissions to compare with by deploying bots running on virtual private server systems (VPS) in the relevant regions.

We chose the scraping audit as it allowed for a simple and transparent way to gather data while leveraging the availability and functionality of web browsers and their add-ons. This facilitated distribution and scalability. Crowdsourcing the collection was a way to involve patients and carers alike to support the endeavour. Inclusion of patient and carers had the added benefit of providing results backed by natural browsing profiles and behaviors.

The plugin was designed to operate on the current versions of Mozilla Firefox and Google Chrome. These browsers were selected because they are among the most popular web browsers and cover the largest percentage of potential data donors [64]. By using two major platforms, we could benefit from their infrastructure that allowed easy distribution of our plugin and possibly gave us some legitimacy by being hosted from the official store sites. The basic structure is illustrated in figure 2.

As the study included recruitment of participants/donors⁷ impacted by the aforementioned diseases, we made a conscious effort to better understand the demographics roughly represented by patients impacted. Specifically, we noted that the average age of diagnosis for Parkinson’s Disease (60+) might present challenges for potential participants/donors (including limited technological literacy and willingness to engage with complicated software) [65]. Thus, the delivered software had to be accessible for this specific demographic.

⁶Diakopoulos denotes *Reverse Engineering* as “the process of articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works” [24, p. 16]

⁷Herein, the notions of *participants* and *donors* are distinct. Participants describe users that only downloaded, installed the plugin and registered, whereas donors are active contributors who submitted their respective collected data.

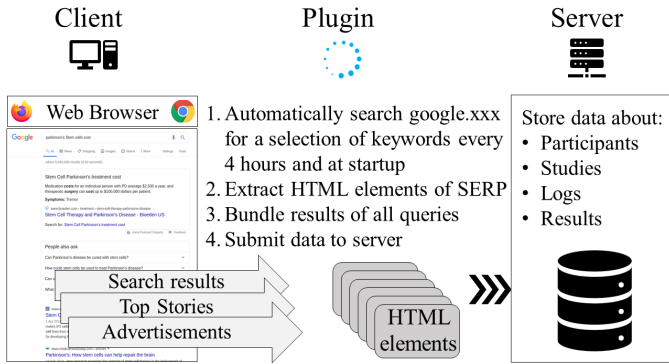


Fig. 2. Sketch of the plugin-server-communication of the EuroStemCell Data Donation, by author

For our purposes, accessibility required a seamless onboarding process and automatic execution with minimal user involvement. Hence, we minimized the number of steps in the registration process, provided FAQs and automatized the donations.

After installation a privacy statement (See Fig 6 in the Appendix) had to be accepted and a minimal questionnaire had to be filled to assign users to study groups and gather some statistical data. Groups were allocated server-side by country and disease with additional control groups for each disease which were subsequently populated by unaffected participants. This ensured that the users were not scattered among the studies and guaranteed the provision of at least one comparative study group.

Once this setup was completed, the plugin automatically sent queries to Google search in a 4-hour cycle starting at midnight and additionally at browser startup. The plugin requested the website `https://www.google.[top level]/search?q=[term]`, where [top level] corresponds to the respective top level domain of a participant group’s region and [term] relates to the search terms that were used in randomized order. Search terms were composed of either a [disease] prefix (“parkinson’s”, “multiple sclerosis”, “diabetes”) followed by clinical terms or “stem cells” in a more general wording. Only the results to these fabricated queries were analyzed, user’s other search activities remained untouched. Then results, ads and top stories on the SERP were collected and sent to the collection server.

All of this happened without user-interaction in a background window to not interfere with a participant’s internet activities. It was part of our effort to provide a balanced user experience between unobtrusive data donation and transparent study design. To support transparency and let users acknowledge their part in the study, the plugin included a utility that showed the user an overview of recent submissions.

It must be emphasised that this study was restricted to the collection of information provided by participants via the initial survey and the data collected by the plugin. Our Black Box analysis cannot be equated with a fully observable In-Out-Relationship in which all input information is visible. Indeed,

this type of one-to-one study of results would require an API that serves as a single source of input [24]. Even then, an opaque system like Google Search may use more than that input. Thus, in our study, the variety of input variables that the algorithms takes into account remain mostly unknown and uncontrollable.

Simply put, the detailed specifics of an algorithm cannot be determined by observers outside of the Black Box. Eventually, the modalities do not need to be known in their completeness to infer about an algorithm’s workings and effects in practice [24]. It is sufficient to “develop a critical understanding of the mechanisms and operational logic” [66, 86]. Rather, the examination should be conducted with focus on relevant aspects only and consider those conditions that are required to understand a phenomenon [67]. Hence, the Black Box analysis of the web-advertisement algorithms of Google conducted can be restricted to the question of whether there still appears questionable advertisements delivered via Google Ads after the announced policy change.

IV. FINDINGS

As this project is ongoing, the following findings reflect our initial analysis of the collected data and indicate areas of developing analysis. Although data was generated for all conditions (Parkinson’s Disease, Multiple Sclerosis, and Diabetes), this analysis will focus on data generated by the Parkinson’s study’s across all geographic locations. Our Parkinson’s Disease studies generated the largest accumulation of both “real” and virtual donors from which we have drawn our findings. Work to dissect both the advertisements and results for Multiple Sclerosis and Diabetes are ongoing.

The Parkinson Disease’s donation campaign operated over the period of September, 30th 2019 to February, 11th 2020. During this period, 162 participants registered (including 24 VPS donors), 102 of which were actually submitting data over the course of the study and contributed 177,756 individual donations. The majority of “real” donors self-identified as being either a patient or carer impacted by Parkinson’s Disease. Additionally, of these “real” donors, the strongest representation of patients and was located in the United Kingdom⁸. VPS servers were located in the UK, Canada, Australia and the United States as a baseline of results across conditions and countries.

Figure 3 shows that we received regular daily donations on a stable level from mid-November on. The VPS-servers worked as expected, submitting in a recurring manner, as depicted by the orange bars. As most “real” donations were submitted between the 4-hour intervals, triggered by a browser startup, this functionality was a vital move for our data collection. This allowed us to capture data even when users were just briefly browsing the web. Although the VPS’ submission frequencies may vary as we see in Figure 3, they were a reliable source of donations.

⁸This may have been influenced by the support of Parkinson’s UK, the Edinburgh Research Interest Group, and the Anne Rowling Clinic located in Edinburgh, Scotland.

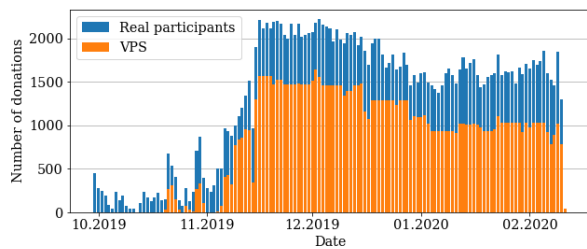


Fig. 3. Donations over time show a spike after onboarding events and VPS rollout

The small number of participants is most likely due to the fact that the average age of patients with a diagnosis of Parkinson’s Disease is relatively high at 60-67 years [65]. This leads to the assumption that most of those affected are not digital natives who use the internet and the google search engine much less frequently to the extent necessary to obtain a sufficiently large sample for a clear conclusion. This study is a good example of the difficulty of persuading affected persons in such a sensitive field of application to participate in a data collection. Although multiple leading research centres joined in collaboration, it shows one of the inherent limitations of a crowd source audit.

A. Presence of Ads Post-Google Moratorium

Between September 2019 and February 2020 we captured 21,188 single advertisements within the data donations. Of these single advertisements, we were able to extract around 285 hosts accounting for paid positioning on the SERPs. The study captured the continued use of targeted advertising through Google’s ad exchange platform for searches around stem cell treatments on Google Search after their self-imposed October 1st deadline. From that date on, promotion of emerging stem cell related practices was, per Google’s public statements, prohibited. As the data captured before and after the policy change contained questionable ads we must conclude that the measures Google introduced are not sufficiently addressing the issue of queries around stem cell treatments. Targeted advertising for stem cell treatment related searches remained continuously present throughout the course of the five-month study. This is particularly worrying as it indicates a break from the stated intentions of Google to prevent the practice due to the potential risk posed to patients of unproven stem cell treatments. Importantly, the rate of stem cell related advertisements appearing in SERPs did not decrease over the period of the study. This indicates that if changes were implemented in the algorithmic behaviour on Google’s side, it did not result in the absence or decrease of advertisements.

B. Captured Ads Analysis

In order to better understand the character of the ads presented, we conducted a structured analysis of these adver-

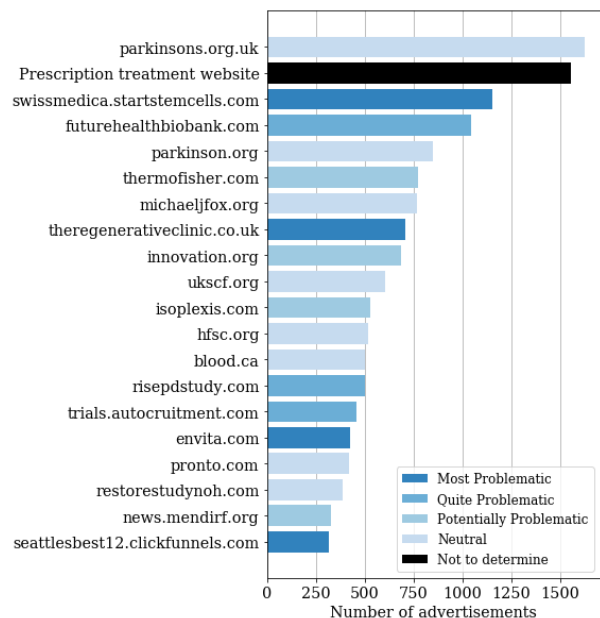


Fig. 4. Top 20 advertisers by ad count.

tisements with the support of EuroStemCell⁹, an EU-funded public engagement project comprised of stem cell researchers and educators from around Europe. This collaboration allowed us to code the sources into a larger constellation within the biomedical industry, public health, and commercial medical services sectors. We then applied a grounded “stoplight” analysis to indicate the potential for risk of users [68] based on the potential for exposure to personal harm represented by both the promotion of unproven stem cell treatments (e.g. unproven mesenchymal stromal cell-derived treatments) and commercial for-profit marketing of treatments.

This analysis showed that the captured ads included a number of potentially problematic sources. These included private clinics advertising direct-to-patient marketing of unproven treatments (including the aforementioned Swiss Medica), for-profit health news websites, private clinical trials, pharmaceutical direct-to-consumer websites, and private biobanking. Less problematic sources included a large number of non-profit and patient advocacy organisations including Parkinson’s UK and the Michael J. Fox Foundation. The Top 20 advertisers included a mix of these sources (see 4).

The second largest source of advertisements captured (labeled ‘Prescription Treatment Website’) sparked multiple redirects. These compiled advertisements only appeared in donations from the United States and did not appear to the plugin as direct links like the other advertisements. Rather, they included obfuscating referring links which appeared to be generated by Google AdSense.

This collection of Prescription Treatment grouped ads were marked by two unique characteristics. The first is that these advertisements appear to have been generated through a different

⁹www.eurostemcell.org

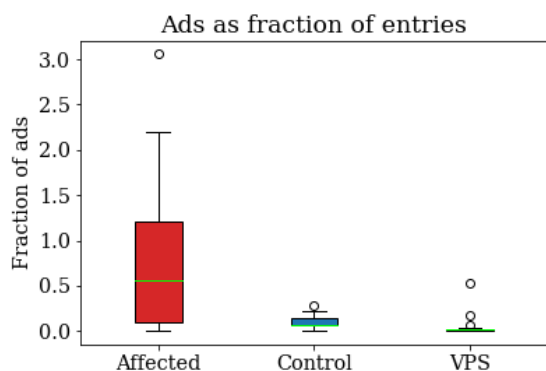


Fig. 5. Fraction of ads in total donations per participant
Means: Affected 0.76, Control 0.10, VPS 0.05
Medians: Affected 0.57, Control 0.07, VPS 0.01

target modality than the other top 20 advertising sources. Each of these links captured by the data donation plugin recorded obfuscated, redirecting links as opposed to clear, direct links. This implies that searches around stem cell treatments and cures and their associated conditions are still included within Google’s algorithmically-driven ad auction ecosystem.

The second characteristic is the promotion of direct-to-consumer marketing of pharmaceuticals for the treatment of conditions (including proprietary versions of Levodopa, a primary pharmacological treatment for the Parkinson’s Disease). These sites were targeted both at medical providers and patients themselves and offered as either direct purchase options or information for patients to supply to their primary care providers. This approach of direct-to-patient advertisement is common in the private healthcare system of United States, but is banned in United Kingdom, Canada or Australia, countries in which health infrastructure is largely publicly governed [69].

C. Ad Targeting of Affected Users

Within the Parkinson’s Disease study groups, data donors who self-identified as either a patient or carer received more advertisements than participants assigned to the control groups or VPS data donors (see Fig. 5).

This indicates that there may be additional modalities within Google’s ad targeting enacting upon users who have been identified as impacted by diseases like Parkinson’s. It is worth noting that the content of these advertisements did not include a larger percentage of highly problematic sources (e.g. private clinics with direct marketing of unproven stem cell treatments). Regardless of the content of the advertisements shown. Specifically to patients and carers, the increased display of advertisement demonstrates that Google’s algorithmic advertising modalities may have already identified vulnerable groups as a target audience.

How and *why* these modalities became visible in our data could not be determined, given the limitations of the study and the group of participants. Further research is urgently needed

to explore this trend and its wider implications for potentially vulnerable users.

V. DISCUSSION

This project is part of a larger interrogation of algorithmic mediation of knowledge in the public health sector and yet it has farther reaching implications. Our initial findings point towards complex questions of responsibility within the tech, biomedical research and healthcare sectors in finding clear boundaries of ethics, commercial interests, cutting-edge medical research and patient safety. These findings open up a host of new questions that must be addressed by many different fields, including algorithmic accountability and governance, tech and bioethics, healthcare decision-making, and patient/carer advocacy.

Although this study provides important evidence of the continued existence of problematic advertisements related to stem cell treatment queries, we are limited in our ability to make concrete statements on *how* and *why* Google’s targeting modalities continue to promote them. Indeed, the nature of Google’s black box platform resists monitoring as part of its proprietary advantage. Therefore, in such sensitive fields of application as the advertising of stem cell therapies, society must demand transparency and accountability verified beyond self-governance undertaken by companies like Google themselves. This includes a shift away from the “repair service behaviour” conducted by Google in the face of evidence of risk to search users. If identification of vulnerable populations is made possible through the algorithmic modalities of user profiling, it should be an imperative to root out these functions as fundamental design flaws rather than commercial features.

We intend to continue our investigation into the data generated by this study with a particular focus on Google as an algorithmic mediator of healthcare and medical decision-making. We intend to tap into the data generated in order to interrogate a number of areas of interest including; quality of SERP results (both advertisement and organic results) across geographical and health infrastructural boundaries, the prevalence of non-profit patient advocacy outreach strategies that feature SEM, and the question of how Google’s SERP ecosystem impacts the work of public health infrastructure and non-profit sector vis-a-vis their relationships to commercial actors. We welcome input from multi-disciplinary researchers interested in better understanding this complex, often-bypassed, yet fundamentally personal touch-point between human health and algorithmically-driven technology.

ACKNOWLEDGMENT

This project has been partially funded by the EU stem cell public engagement project, EuroStemCell (www.eurostemcell.org) and by a generous grant from the University of Edinburgh School of Social and Political Science. The research was associated with the project GOAL “Governance of and by algorithms” (Funding code 01IS19020) which is funded by the German Federal Ministry of Education and Research.

ETHICS

This project and the associated methodological plug-in was reviewed and approved through University of Edinburgh School of Social and Political Science (Approved 29.03.2019 in association with the doctoral research of Anna Couturier "Google Search and the Mediation of Knowledge on Treatments: A Case Study on Unproven Stem Cell Treatments" [UoE Internal ID: 259679]). Additionally, we consulted with Prof. Clare Blackburn (School of Biological Sciences, University of Edinburgh, MRC Centre for Regenerative Medicine) and patient advocates (Alison Williams - Parkinson's UK) throughout the process. All efforts have been made to ensure that study participants maintained anonymity and informed consent.

REFERENCES

- [1] Food and Drug Administration, "FDA Warns About Stem Cell Therapies," 2019. [Online]. Available: <https://www.fda.gov/consumers/consumer-updates/fda-warns-about-stem-cell-therapies>
- [2] Eurostemcell, "What diseases and conditions can be treated with stem cells?" 2020. [Online]. Available: <https://www.eurostemcell.org/what-diseases-and-conditions-can-be-treated-stem-cells>
- [3] ISSCR, "How to Report False Marketing Claims and Adverse Events from Clinics Offering Unapproved Stem Cell "Therapies";" 2019. [Online]. Available: <https://www.closerlookatstemcells.org/patient-resources/how-to-report-false-marketing-claims-and-adverse-events-from-clinics-offering-unapproved-stem-cell-therapies>
- [4] A. Petersen, C. Tanner, and M. Munsie, "Citizens' use of digital media to connect with health care: Socio-ethical and regulatory implications," *Health (London, England : 1997)*, vol. 23, no. 4, pp. 367–384, 2019.
- [5] C. Tanner, M. Munsie, D. Sipp, L. Turner, and C. Wheatland, "The politics of evidence in online illness narratives: An analysis of crowdfunding for purported stem cell treatments," *Health (London, England : 1997)*, vol. 23, no. 4, pp. 436–457, 2019.
- [6] L. Turner, "ClinicalTrials.gov, stem cells and 'pay-to-participate' clinical studies," *Regenerative medicine*, vol. 12, no. 6, pp. 705–719, 2017.
- [7] S. Yuan, A. Z. Abidin, M. Sloan, and J. Wang, "Internet advertising: An interplay among advertisers, online publishers, ad exchanges and web users," *arXiv preprint arXiv:1206.1754*, 2012. [Online]. Available: <https://arxiv.org/pdf/1206.1754.pdf>
- [8] Z. Master, K. Robertson, D. Frederick, C. Rachul, and T. Caulfield, "Stem cell tourism and public education: the missing elements," *Cell stem cell*, vol. 15, no. 3, pp. 267–270, 2014.
- [9] S. Rashtchy, A. M. Kessler, P. J. Bieber, and Schindler, Nathaniel H., Tzeng, Judith C., "The User Revolution: The New Advertising Ecosystem And The Rise Of The Internet As A Mass Medium." [Online]. Available: <http://people.ischool.berkeley.edu/~hal/Courses/StratTech09/Lectures/Google/Articles/user-revolution.pdf>
- [10] S. Muthukrishnan, "Bidding on Configurations in Internet Ad Auctions," in *Computing and Combinatorics*, H. Q. Ngo, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1–6.
- [11] Google, "About automation with Google Ads," 2019. [Online]. Available: https://support.google.com/google-ads/answer/9297584?hl=en&ref_topic=6294205
- [12] —, "Determine a bid strategy based on your goals," 2020. [Online]. Available: <https://support.google.com/google-ads/answer/2472725>
- [13] —, "Personalized advertising," 2020. [Online]. Available: <https://support.google.com/adspolicy/answer/143465>
- [14] —, "List of Ad policies: Healthcare and medicines," 2019. [Online]. Available: <https://support.google.com/adspolicy/answer/176031>
- [15] A. Biddings, "Google Ads Help: A new policy on advertising for speculative and experimental medical treatments," 2019. [Online]. Available: <https://support.google.com/google-ads/answer/9475042?hl=en>
- [16] W. Wan and L. McGinley, "New Google policy bars ads for unproven stem cell therapies." [Online]. Available: <https://www.washingtonpost.com/health/2019/09/06/new-google-policy-bars-ads-unproven-stem-cell-therapies/>
- [17] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019.
- [18] L. D. Introna, "Algorithms, Governance, and Governmentality," *Science, Technology, & Human Values*, vol. 41, no. 1, pp. 17–49, 2016.
- [19] N. Seaver, "Knowing Algorithms," in *Media in Transition*, Cambridge, MA, 2014, vol. 8. [Online]. Available: https://digitalsts.net/wp-content/uploads/2019/03/26_Knowing-Algorithms.pdf
- [20] T. Gillespie, "The Relevance of Algorithms," *Media technologies: Essays on communication, materiality, and society*, vol. 167, 2014.
- [21] J. Bennett, *Vibrant matter: A political ecology of things*. Duke University Press, 2010.
- [22] D. Pedreschi, F. Giannotti, R. Guidotti, A. Monreale, L. Pappalardo, S. Ruggieri, and F. Turini, "Open the Black Box Data-Driven Explanation of Black Box Decision Systems." [Online]. Available: <http://arxiv.org/pdf/1806.09936v1>
- [23] R. Kitchin, "Thinking critically about and researching algorithms," *Information, Communication & Society*, vol. 20, no. 1, pp. 14–29, 2017.
- [24] N. Diakopoulos, "Algorithmic Accountability Reporting: On the Investigation of Black Boxes." [Online]. Available: <https://academiccommons.columbia.edu/doi/10.7916/D8TT536K/download>
- [25] D. Dörner, *Die Logik des Mißlingens: Strategisches Denken in komplexen Situationen*. Rowohlt Verlag GmbH, 2011.
- [26] C. Sandvig, K. Hamilton, K. Karahalios, and C. Langbort, "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms," in *Data and Discrimination: Converting Critical Concerns into Productive*, 2014.
- [27] N. Diakopoulos, "Algorithmic Accountability. Journalistic investigation of computational power structures," *Digital Journalism*, no. 3, pp. 398–415, 2015. [Online]. Available: <https://doi.org/10.1080/21670811.2014.976411>
- [28] USACM, "Statement on Algorithmic Transparency and Accountability." [Online]. Available: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
- [29] S. Erikainen, A. Couturier, and S. Chan, "Marketing experimental stem cell therapies in the uk: biomedical lifestyle products and the promise of regenerative medicine in the digital era," *Science as Culture*, vol. 29, no. 2, pp. 219–244, 2020.
- [30] Eurostemcell. (2020) What diseases and conditions can be treated with stem cells? [Online]. Available: <https://www.eurostemcell.org/what-diseases-and-conditions-can-be-treated-stem-cells>
- [31] T. Caulfield, D. Sipp, C. E. Murry, G. Q. Daley, and J. Kimmelman, "Confronting stem cell hype," *Science*, vol. 352, no. 6287, pp. 776–777, 2016.
- [32] A. Petersen, M. Munsie, C. Tanner, C. MacGregor, and J. Brophy, *Stem cell tourism and the political economy of hope*. Springer, 2017.
- [33] C. MacGregor, A. Petersen, and M. Munsie, "From the margins to mainstream: How providers of autologous 'stem cell treatments' legitimise their practice in Australia," *Health*, 2019. [Online]. Available: <https://doi.org/10.1177/1363459319846927>
- [34] C. B. Cohen and P. J. Cohen, "International stem cell tourism and the need for effective regulation: Part i: Stem cell tourism in russia and india: clinical research, innovative treatment, or unproven hype?" *Kennedy Institute of Ethics Journal*, vol. 20, no. 1, pp. 27–49, 2010.
- [35] T. Lysaght, W. Lipworth, T. Hendl, I. Kerridge, T.-L. Lee, M. Munsie, C. Waldby, and C. Stewart, "The deadly business of an unregulated global stem cell industry," *Journal of medical ethics*, vol. 43, no. 11, pp. 744–746, 2017.
- [36] D. Sipp, T. Caulfield, J. Kaye, J. Barfoot, C. Blackburn, S. Chan *et al.*, "Marketing of unproven stem cell-based interventions: A call to action. sci," *Trans. Med*, vol. 9, 2017.
- [37] A. C. Regenber, L. A. Hutchinson, B. Schanker, and D. J. Mathews, "Medicine on the fringe: stem cell-based interventions in advance of evidence," *Stem Cells*, vol. 27, no. 9, pp. 2312–2319, 2009.
- [38] A. K. McLean, C. Stewart, and I. Kerridge, "Untested, unproven, and unethical: the promotion and provision of autologous stem cell therapies in australia," *Stem cell research & therapy*, vol. 6, no. 1, pp. 1–8, 2015.
- [39] R. Mendick and A. Hall, "Europe's largest stem cell clinic shut down after death of baby," *The telegraph*, vol. 8, 2011.
- [40] Z. Tufekci, "Engineering the public: Big data, surveillance and computational politics," *firstmonday*, vol. 19, no. 7, 2014. [Online]. Available: <https://firstmonday.org/ojs/index.php/fm/article/view/4901>
- [41] B. Krishnamurthy and C. E. Wills, "Privacy Diffusion on the Web: A Longitudinal Perspective," in *Proceedings of the 18th international*

- conference on World Wide Web. Madrid: ACM, 2009, pp. 541–550. [Online]. Available: <http://www2009.eprints.org/551/p541.pdf>
- [42] S. Goel, J. M. Hofman, and M. I. Sirovica, “Who Does What on the Web: A Large-Scale Study of Browsing Behavior,” *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.
- [43] G. Ren, E. Tune, T. Moseley, Y. Shi, S. Rus, and R. Hundt, “Google-wide profiling: A continuous profiling infrastructure for data centers,” *IEEE micro*, vol. 30, no. 4, pp. 65–79, 2010.
- [44] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large datasets (how to break anonymity of the Netflix prize dataset),” *University of Texas at Austin*, 2008.
- [45] M. Van Otterlo, “A machine learning view on profiling,” *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology*. Abingdon: Routledge, pp. 41–64, 2013.
- [46] J. Angwin, M. Varner, and A. Tobin, “Facebook Enabled Advertisers to Reach ‘Jew Haters’,” 2017. [Online]. Available: <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>
- [47] J. Angwin and T. Parris Jr., “Facebook Lets Advertisers Exclude Users by Race,” 2016. [Online]. Available: <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>
- [48] A. Datta, M. C. Tschantz, and A. Datta, “Automated experiments on ad privacy settings,” *Proceedings on Privacy Enhancing Technologies*, vol. 2015, no. 1, pp. 92–112, 2015. [Online]. Available: <https://www.andrew.cmu.edu/user/danupam/dtd-pets15.pdf>
- [49] Google, “Where your ads will appear on Google,” 2019. [Online]. Available: <https://support.google.com/google-ads/answer/6335981>
- [50] —, “Personalized Search Graduates from Google Labs,” 2019. [Online]. Available: http://googlepress.blogspot.com/2005/11/personalized-search-graduates-from_10.html
- [51] E. Steel and J. Angwin, “On the web’s cutting edge, anonymity in name only,” *The Wall Street Journal*, vol. 4, 2010. [Online]. Available: <https://www.wsj.com/articles/SB10001424052748703294904575385532109190198>
- [52] S. Guha, B. Cheng, and P. Francis, “Challenges in measuring online advertising systems,” in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, 2010, pp. 81–87.
- [53] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen, “How much can behavioral targeting help online advertising?” in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 261–270. [Online]. Available: <https://dl.acm.org/doi/10.1145/1526709.1526745>
- [54] S. C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell, “Psychological targeting as an effective approach to digital mass persuasion,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 48, pp. 12714–12719, 2017.
- [55] M. Lawson. Grow your business faster with machine learning: Part i. [Online]. Available: <https://www.blog.google/products/ads/adwords-machine-learning-part-1>
- [56] D. Mitby. Grow your business faster with machine learning: Part ii. [Online]. Available: <https://www.blog.google/products/ads/adwords-machine-learning-part-2>
- [57] B. Reese. Grow your business faster with machine learning: Part iii. [Online]. Available: <https://www.blog.google/products/ads/adwords-machine-learning-part-3>
- [58] M. Chow. Ai and machine learning get us one step closer to relevance at scale. [Online]. Available: <https://www.thinkwithgoogle.com/future-of-marketing/emerging-technology/ai-personalized-marketing/>
- [59] Google, “Misrepresentation,” 2019. [Online]. Available: <https://support.google.com/adspolicy/answer/6020955>
- [60] A. Goldfarb and C. Tucker, “Advertising bans and the substitutability of online and offline advertising,” *Journal of Marketing Research*, vol. 48, no. 2, pp. 207–227, 2011.
- [61] W. R. Ashby, *An Introduction to Cybernetics*. London: Chapman & Hall, 1957.
- [62] C. Sandvig, K. Hamilton, K. Karahalios, and C. Langbort, “Auditing algorithms: Research methods for detecting discrimination on internet platforms,” *Data and discrimination: converting critical concerns into productive inquiry*, vol. 22, 2014.
- [63] T. D. Krafft, M. P. Hauer, and K. A. Zweig, “Why do we need to be bots? what prevents society from detecting biases in recommendation systems,” in *International Workshop on Algorithmic Bias in Search and Recommendation*. Springer, 2020, pp. 27–34.
- [64] statcounter, “Browser Market Share Worldwide,” 2019. [Online]. Available: <https://gs.statcounter.com/browser-market-share>
- [65] G. Pagano, N. Ferrara, D. J. Brooks, and N. Pavese, “Age at onset and parkinson disease phenotype,” *Neurology*, vol. 86, no. 15, pp. 1400–1407, 2016.
- [66] T. Bucher, “Neither Black Nor Box: Ways of Knowing Algorithms,” in *Innovative Methods in Media and Communication Research*, S. Kubitschko and A. Kaun, Eds. Cham: Springer International Publishing, 2016, pp. 81–98.
- [67] A. Grunwald, *Technikfolgenabschätzung - Eine Einführung*. Berlin: edition sigma, 3-89404-931-6, 2002.
- [68] B. G. Glaser et al., “Strauss (1967): The discovery of grounded theory: Strategies for qualitative research,” *London: Wiedenfeld and Nicholson*, vol. 81, p. 86, 1978.
- [69] C. L. Ventola, “Direct-to-consumer pharmaceutical advertising: therapeutic or toxic?” *Pharmacy and Therapeutics*, vol. 36, no. 10, p. 669, 2011.

APPENDIX

Privacy Statement

The purpose of this statement is to ensure that you have read and understood the information about the study and are fully aware of your rights should you decide to take part. If you would like to take part, please indicate this by reading the following questions. Consent is required in order to download and install this plug-in.

Please note: since data is anonymised at the point of contact, we cannot retroactively withdraw any data collected before uninstallation.

Declaration of consent:

With the installation of the plugin I confirm that the first 10 search results and ads of Google on Google.com (in "All") including the above mentioned additional data (plugin ID, time, exact query, approximate location) on my browser will be made available to the public under a CC-0 license for analysis. I understand that the plugin regularly searches for all healthcare related queries listed above.

I agree that the following data will be collected, processed and published:

- A general location, derived from the IP address, corresponding in precision to approximately your postal code.
- The plug-in ID
- The exact search query
- The time of the search
- The result of the search (everything on the first page of the search results)
- The language setting of the browser
- Whether you are logged in as a user with Google.

I can disable or de-install the plugin at any time. No further data is sent with the deactivation of the plug-in or with its de-installation.

By downloading the plugin, I confirm that I have read and understand the privacy statement, www.eurostemcell.org/datadonation, for the above study. I have had the opportunity to consider the information and ask questions which have been answered satisfactorily

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

I understand that any personal data collected during the study will be treated with confidence and handled in accordance with the Data Protection Act 1998 and GDPR.

I agree to take part in the above study.

Fig. 6. Screenshot of the Privacy Statement of the Eurostemcell Data Donation Campaign, 05.08.2020, from <https://www.Eurostemcell.org/datadonation#paragraph-1576>